

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Molecular interaction fields based descriptors to interpret and compare chromatographic indexes

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/121445> since

Published version:

DOI:10.1016/j.chroma.2012.06.069

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



UNIVERSITÀ DEGLI STUDI DI TORINO

This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process - such as editing, corrections, structural formatting, and other quality control mechanisms - may not be reflected in this version of the text. The definitive version of the text was subsequently published in.

Journal of Chromatography A
Volume 1252, 24 August 2012, Pages 84–89
DOI: 10.1016/j.chroma.2012.06.069

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

(1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.

(2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.

(3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>)

<http://www.sciencedirect.com/science/article/pii/S0021967312009636>

24 Molecular Interaction Fields based descriptors to
25 interpret and compare chromatographic indexes

26 *G. Ermondi and G. Caron**

27 CASMedChem research group, DSTF at the Centre for Innovation, Università degli Studi di Torino, via
28 Quarello 11, 10135 Torino, Italy.

29
30 E-mail: giulia.caron@unito.it

31 Telephone: +39 011 6708337

34 **Abstract**

35 Molecular Interaction Fields (MIFs) based descriptors can be conveniently used to characterize and
36 compare chromatographic scales. In this study, Quantitative Structure–Retention Relationships (QSRR)
37 for eight different chromatographic systems were obtained with VolSurf+ descriptors and Partial Least
38 Squares (PLS). A new and purpose-designed analysis tool highlights the different balance of
39 intermolecular interactions governing solute retention, and estimates the similarity between
40 chromatographic systems.

41

42

1. Introduction

Chromatographic indexes determined by HPLC have recently gained considerable relevance in Absorption, Distribution, Metabolism and Excretion (ADME) studies, because they are faster and easier to obtain than log P (P is the partition coefficient in octanol/water system), and can provide significant information about the compounds' lipophilicity, with or without conversion to log P [1-4]. The large variety of stationary phases and mobile-phase additives that are now available provides the flexibility enabling chromatography to be used to adjust the properties of the partitioning phases [5], so as to directly model biological partition rather than mimicking it. However, to model biological partition with chromatographic data, it must be verified that the balance of intermolecular forces governing the two systems is the same.

In more general terms, the deconvolution of the intermolecular interactions governing chromatographic retention is a topic of great interest in many analytical fields. Modeling tools are routinely used for this purpose [2, 6]. The linear solvation energy relationship (LSER) is currently the most widely-used approach to understanding the types and relative strengths of the chemical interactions controlling retention and selectivity. One of the most generally accepted symbolic representations of the LSER model is given in Eq. 1 [7]

$$\log k = c + eE + sS + aA + bB + vV \quad \text{Eq. 1}$$

The model consists of product terms representing solute properties (descriptors), indicated by capital letters, and the complementary system properties, indicated by lower-case letters. Each product term describes the relative contribution of specific intermolecular interactions to the correlated solute property, in this case, log k (k is the retention factor). The contribution from electron lone pair interactions is expressed by eE, interactions of a dipole type by sS, hydrogen-bond interactions by aA and bB, and differences in cavity formation and dispersion interactions for transfer of the solute from one phase to the other, by vV. The solute descriptors (often called solvatochromic descriptors) are

69 formally defined as excess molar refraction, E; dipolarity/polarizability, S; effective hydrogen-bond
70 acidity, A; effective hydrogen-bond basicity, B; and McGowan's characteristic volume, V.
71 Solvatochromic descriptors are empirical in character, and thus only available for a limited number of
72 compounds. In general, solvatochromic descriptors are lacking for multiple functional and for ionised
73 compounds [2, 8-10].
74 For this reason it appeared valuable to replace solvatochromic descriptors with more convenient
75 descriptors of a fully computational nature, based solely on the 3D structural formula of compounds, i.e.
76 VolSurf+ descriptors. These are based on 3D molecular fields [11, 12] and so could lead to a rational
77 explanation of retention differences in terms of intermolecular interactions. Briefly, a 3D molecular
78 field may be viewed as a 3D matrix, with attractive and repulsive energy values between a chemical
79 probe and a target molecule; VolSurf+ is a tool that extracts molecular descriptors from the 3D matrix.
80 Figure 1 shows the visual representation of some VolSurf+ descriptors calculated for amitrole (example
81 of target molecule present in the dataset of the investigated compounds).

82

83 *Insert Figure 1.*

84

85 Figure 1A shows the envelope corresponding to three VolSurf+ descriptors (W3, W5 and W7) obtained
86 with the water probe. They represent the volume of the molecular envelope which is accessible to, and
87 interacts attractively with water molecules at three levels of interaction energy (-1, -3 and -5kcal/mol).
88 Since at different energy values the contribution of the intermolecular (and intramolecular-) interactions
89 to the single descriptor varies, VolSurf+ descriptors are better interpreted using blocks of descriptors
90 obtained with a single probe.

91 VolSurf descriptors can therefore be classified in six blocks (see Supporting Information): a) descriptors
92 that characterize the size and shape of the solute (7 descriptors in the text, briefly called size, color-code
93 green), b) 19 descriptors that express the solute's interaction with water molecules (in the text indicated
94 as water, color-code light blue), c) 5 descriptors that describe the solute's ability to form hydrogen bond

interactions with the donor group of the probe (that mimics the chromatographic system, color-code blue, see below), d) 5 descriptors expressing the solute's ability to form hydrogen bond interactions with the acceptor group of the probe (that mimics the chromatographic system, color-code red, see below) e) 28 descriptors describing the solute's propensity of the solute to participate in hydrophobic interactions (in the text called DRY for short, color-code yellow), f) 18 descriptors mainly describing the imbalance between hydrophilic and hydrophobic regions (in the text called others, color-code grey). Please note that for the sake of clarity, to identify hydrogen bonding (HB) interactions (Hydrogen Bond Acceptor capability (HBA)/Hydrogen Bond Donor capability (HBD) and color code) we refer to the probe's properties and not to the solute. This is shown in Figure 1C.

Some years ago, VolSurf descriptors and PLS were used to calculate and interpret log P in alkane/water system [13]. In 2011 the same strategy was applied to model log k₃₀ (k₃₀ is the isocratic retention factor) data for a series of Pt(II) derivatives [14].

It is here shown that MIFs-based descriptors can conveniently be used to quali- and quantitatively characterize any chromatographic index on the basis of the intermolecular interactions governing the system. For this purpose, QSRR models based on PLS algorithms were firstly made. The study's end-point was to develop graphical and numerical tools at different levels of complexity that enable researchers to extract as much chemical information as possible from a series of chromatographic systems, and numerically rank their similarity/dissimilarity.

2. Methods

Frog 2.1 (<http://bioserv.rpbs.univ-paris-diderot.fr/cgi-bin/Frog2>) was used to build the 3D structures that were then submitted to AM1 minimization as implemented in Spartan '08 (Wavefunction Inc, Irvine, CA). For any compound, the lowest energy conformation as obtained from Spartan minimization was selected as the final structure, saved in mol2 format and submitted to VolSurf+ (version 1.0.4, Molecular Discovery Ltd. Pinner, Middlesex, UK, 2009, <http://www.moldiscovery.com>) using default

120 settings and four probes (OH2, DRY N1 and O probes that mimic respectively water, hydrophobic,
121 hydrogen bond acceptor and hydrogen bond donor interaction of the compounds with the environment).
122 PCA and PLS tools implemented in VolSurf+ were used.

123

124 **3. Results and discussion**

125 ***3.1. The chromatographic systems***

126 The dataset used was that developed by Tuel et al. [9], containing 76 structurally-diverse and complex
127 organic structures, chiefly pesticides together with some drugs. The full list of chemical structures is in
128 the Supporting Information of the original paper [9].

129 The capacity factors of the 76 compounds were determined in eight HPLC systems [9]. The
130 specifications of the HPLC systems are in Table 1; further details are in the original paper [9]. Each
131 system is identified by an abbreviation (Table 1), used throughout the text. Six different columns were
132 investigated and different composition of the mobile phases were adopted. Log k values were available
133 for four reversed phase (RP), three normal phase (NP) and one hydrophilic interaction (Hydrophilic
134 Interaction Liquid Chromatography, HILIC) chromatographic systems (Table S1 in the Supporting
135 Information shows the descriptive statistics of log k). Taken together retention characteristics were
136 highly diverse as confirmed by the correlation matrix (Table S2 in the Supporting Information) that
137 evidences a large variability in cross correlations (Pearson coefficients).

138 ***3.2. Building PLS models***

139 To relate log k values with the chemical structures of compounds mathematically, a PLS study was
140 performed as will be described.

141 Firstly, the 3D structures of the 76 molecules in the dataset were built (see Experimental Part) and, for
142 each structure, 82 descriptors (the full list of descriptors is available in Table S3 of the Supplementary
143 Information) were calculated by VolSurf+.

Log k values for the eight chromatographic systems [9] were then imported into VolSurf+ as dependent variables (Y) and a relation between Y and VolSurf+ descriptors (X) was sought using the standard PLS tool implemented in the software. The statistics of the best models are shown in Table 1.

Insert Table 1

To validate these models internal validation was firstly used. Whereas some researchers in the QSAR field support internal validation, others consider that internal validation is not a sufficient test to check the robustness of models, and external validation is necessary [15, 16]. In this case, however, since the sample size is small, and thus holding a portion of it back for testing would be wasteful, it was preferred to use cross-validation, with multiple rounds using different partitions (Table 1 shows Q^2 for the LOO procedure, but results were similar with different partitions). Most of the PLS models in Table 1 (apart from the CN-NP system, which is a limit situation) showed $R^2 > 0.6$ and $Q^2 > 0.5$, and satisfied the Tropsha et al. validation rules [17]. Finally, to further validate PLS models we randomized the order of Y values which produced unacceptable R^2 and Q^2 values (data not shown).

The relatively low R^2 values might be due to the complexity and molecular flexibility of the dataset, whereas the small difference between R^2 and Q^2 supports the stability of the models. Some outliers were identified for 6 out of 8 systems. An in-depth analysis of the reasons for the presence of outliers (most of them probably related to the flexibility of compounds) is outside the scope of this study, but it is worth noting that the inclusion/exclusion of outliers slightly altered PLS statistics, but did not alter the chemical interpretation of the PLS models (see below).

3.3. Mechanistic interpretation of PLS models

The main drawback of PLS models is the difficulty of interpreting them from the chemical standpoint. Variable Importance in Projection (VIP) plots are often used to overcome this limit [13]. The VIP represents the value of each predictor in fitting the PLS model; the VIP values in the plot summarise the overall contribution of each X-variable to the PLS model, summed over all components and weighted

according to the Y variation accounted for by each component (Eq. 2). The sum of squared VIP values is equal to the number of descriptors [15].

$$VIP_j = \left\{ \frac{\sum_{f=1}^F w_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F} \right\}^{1/2} \quad Eq. 2$$

Where w_{jf} is the weight value for variable j component f , SSY_f is the sum of squares of explained variance for the f th component and J the number of variables. SSY_{total} is the total sum of squares explained of the dependent variable, and F is the total number of components. The weights in a PLS model reflect the covariance between the independent and dependent variables and the inclusion of the weights is what allows VIP to reflect not only how well the dependent variable is described but also how important that information is for the model of the independent variables. Note that $\sum_{j=1}^F VIP_j^2 = J$. As explained in the Introduction, the best interpretation of VolSurf+ descriptors is achieved when they are grouped into six blocks. Therefore, to obtain a fully optimized interpretation of the PLS models (Table 1) BlockRelevance (BR) was defined as the ratio of the sum of the squared VIP values of a given block of descriptors and the number of those descriptors (Eq. 3).

$$BR_i = (\sum_{j=1}^{N_i} VIP_j^2) / N_i \quad Eq. 3$$

where i is the number of blocks (6 here, Table 2), N is the number of descriptors for any block, VIP is the value of each predictor fitting the PLS model.

BR shows the relevance of a certain block of descriptors to a given chromatographic system; the higher the value of BR, the more important is that block. The set of BR values can therefore be used as a first set of numbers able to characterize a chromatographic system.

193 Table 2 lists the BR values for the 8 systems investigated. For example, C18 is mainly characterized by
194 the size block of descriptors, (BR = 2.35) which is at least twice as important as any other block. A
195 similar trend is found for Ph. The CN-RP system is almost equally characterized by HBD (Fig. 1C), size
196 and water descriptors; the situation is similar for OH-RP. HILIC is mainly characterized by size and
197 HBD. The contribution of water and size descriptors is most significant for OH-NP, whereas HBD
198 properties are the mainly determinant of the NH₂-NP system.

199

200 *Insert Table 2*

201

202 From BR analysis, it is known, for example, that a given block of solutes descriptors characterizes a
203 system (e.g. size for C18, Table 2) but not if the corresponding intermolecular interactions promote the
204 retention either in the stationary or in the mobile phase. The sign of the PLS coefficient is informative: a
205 positive coefficient means that an increase of the descriptor considered causes an increase in log k (and
206 thus the retention in stationary phase increases); the reverse is true for negative coefficients. Depending
207 on the sign of the PLS coefficient, BR was broken down into BR (+) relating to retention in the
208 stationary phase, and BR (-), relating to retention in the mobile phase (Eq. 4).

209

$$210 \quad BR_i = BR_i (+) + BR_i (-) \quad Eq. 4$$

211

212 BR (+) and BR (-) for the eight investigated systems are illustrated in Fig. 2; this is a significant tool to
213 compare chromatographic systems.

214

215 *Insert Figure 2.*

216

217 Figure 2 clearly distinguishes RP from NP systems, as shown by the different trends of the two blocks
218 of VolSurf+ descriptors expressing size and interaction with the water probe. In RP systems, the size

219 block contributes significantly to increasing interaction with the stationary phase, whereas the same
220 block is significant and of opposite sign in NP systems. Interestingly, for HILIC the contribution of the
221 size block is very important, but of negative sign, distinguishing this system from both RP and NP, as
222 might be expected [18]. The block of water descriptors is of negative sign in RP systems, but positive
223 for NP and HILIC systems. This confirms that hydrophilic compounds prefer the mobile phase in RP
224 systems, whereas they have greater affinity for the stationary phase in NP and HILIC systems.

225 The blocks size, HBA and (although in minor extent) HBD enable the four RP systems to be classified.
226 The contribution of the size descriptors varies in the following order: C18~Ph>CN-RP~OH-RP. The
227 HBD trend is similar in the four systems, but the reverse is true for HBA. Ph and OH-RP show opposite
228 trends in the HBA block (red in Figure 2, negative for Ph and positive for OH-RP). This confirms that
229 HBA properties are mainly expressed in the stationary phase of the OH-RP system, and in the mobile
230 phase of Ph system.

231 The three NP systems investigated share similar trends for size, water and dry blocks. The most
232 significant differences are again found for HB-related descriptors. The HBA block (red in Fig. 2) is
233 large and positive for NH₂-NP, about half the size and negative for CN-NP, with OH-NP in an
234 intermediate situation. The HBD block shows the same trend for the three systems, being particularly
235 important for CN-RP. To sum up, in qualitative terms NH₂-NP is different from CN-NP and from OH-
236 NP, which are much closer to one another (see the quantitative analysis below).

237 As expected the HILIC system shows a unique profile. As mentioned above, the block of size
238 descriptors is the most important whereas the block of HBA is slightly less so. These findings show that
239 two polar compounds with similar size can be separated by the HILIC system investigated (Table 1) on
240 the basis of their HBD properties. This supports experimental evidence about the application of HILIC
241 systems containing high amounts of acetonitrile in the mobile phase to the separation of amines and
242 peptides. [18]. It should be stressed that application of this analysis method to different HILIC systems
243 could, significantly, improve a thorough understanding of retention behavior of HILIC, thus enhancing
244 their applicability domain.

OH-RP and OH-NP are a pair of chromatographic systems characterized by the same column but different mobile phase composition (polar for the RP system and apolar for the NP system; see Table 1). As Figure 2 shows, in the presence of the same stationary phase, different mobile phases can completely alter the balance of intermolecular forces responsible for retention.

OH-NP and NH₂-NP are a pair of systems that use the same mobile phase (Table 1) but different stationary phases. Figure 2 highlights the different performances of the two columns in terms of intermolecular interactions, and suggests that this analysis could be particularly important when new columns are launched on the market.

Finally, a further requirement is a numerical value with which to quickly rank the diversity of chromatographic systems. For this purpose, a simple match parameter called LipophilicityMatch (LM) (Eq. 5), was defined

$$LM_{i,j} = \sum_{f=1}^{82} f(c_f^i; c_f^j) \quad Eq. 5$$

where $f(c_f^i; c_f^j) = 1$ for $c_f^i \bullet c_f^j \geq 0$

$= 0$ for $c_f^i \bullet c_f^j < 0$

and c is the coefficient and i and j are a pair of chromatographic systems

LM values for the eight systems investigated are reported in Table 3.

Insert Table 3

As expected, the RP and NP systems have little similarity (not above 33%). Among the RP systems investigated, LM values range from 85% to 76%. Among the NP systems, LM values range from 66 to 77% confirming the qualitative results discussed above. The LM analysis also confirms that the HILIC system is more similar to NP (about 70%) than to RP (about 28%) systems.

3.4. *How to individuate the chromatographic system that best mimics the octanol/water system*

A way of comparing $\log P_{\text{oct}}$ with $\log k$ values could facilitate the implementation of relevant $\log k$ descriptors in running research programs in the pharmaceutical industry. This comparison can easily be determined by LM as described below. For the above mentioned dataset of compounds a good PLS model ($R^2 = 0.82$, $Q^2 = 0.70$, $LV = 3$) was obtained using $\log P_{\text{oct}}$ (data by Tuelp et al. [9]) as dependent variable and VolSurf+ descriptors as independent variables. Figure 3A shows the trend of BR (+) and BR (-) for the octanol/water system. The size block is large and positive. The remaining blocks are less important. LM were calculated and shown in Figure 3B. As expected, the highest values were found for RP systems. Given the LM definition (Eq. 5), a minute comment of the numerical differences should be avoided.

Insert Figure 3.

4. Conclusions

The QSRR analysis performed in this study demonstrates the goodness of the VolSurf+ descriptors, for describing the mechanism of chromatographic retention at the molecular level. The PLS models based on VolSurf+ descriptors were interpreted via custom-designed parameters (BR and LM) to fully exploit the interpretative potential of the VolSurf+ descriptors in quali- and quantitative terms.

The potential applications of this kind of analysis are numerous. In drug discovery programs, the reliability of lipophilicity indexes determined on immobilized biomacromolecule stationary phases could be verified, and biomimetic chromatographic partition systems may prove to be better models than $\log P$ for use in ADME studies, or otherwise. From a more general point of view, manufacturers, for example, could compare new and old products, definitively managing end-capping effects and identifying columns with equivalent properties. In proteomic studies, metabolites could be identified and/or classified by their different behavior on chromatographic systems, revealing their differing balance of intermolecular interactions.

297 **Acknowledgement**

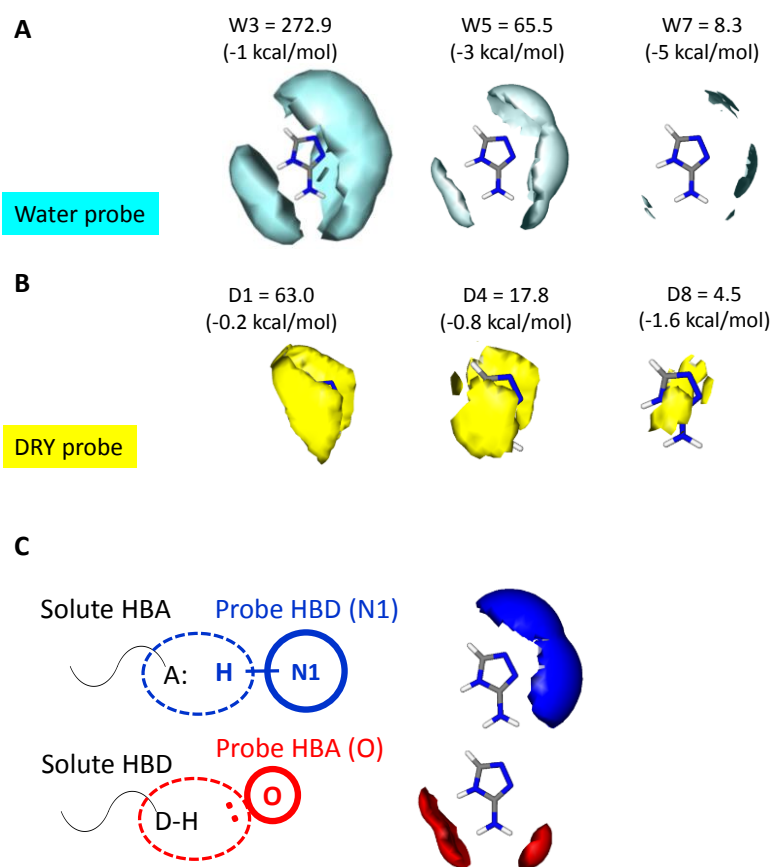
298 GE and GC thank Dr. Fenner for providing additional experimental details to published results

299 **References**

- 300 [1] G. Caron, R.A. Scherrer, G, Ermondi,. in: Testa B., van de Waterbeemd H. (Eds). Comprehensive
301 Medicinal Chemistry, 2nd Edition, Vol. 5, Oxford, UK: Elsevier; 2006. pp. 425.
- 302 [2] R. Kaliszan, Chem. Rev. 107 (2007) 3212.
- 303 [3] K. Valko, J. Chromatogr. A 1037 (2004) 299.
- 304 [4] S. K. Poole and C. F. Poole, J. Chromatogr. B 797 (2003) 3.
- 305 [5] F. Lombardo, M. Y. Shalaeva, K. A. Tupper, F. Gao, and M. H. Abraham, J. Med. Chem. 43
306 (2000) 2922.
- 307 [6] R. Put and Y. Vander Heyden, Anal. Chim. Acta 602 (2007) 164.
- 308 [7] M. H. Abraham, Chem. Soc. Rev. 14 (1993) 153.
- 309 [8] J. Platts, M. Abraham, D. Butina, and A. Hersey. J. Chem. Inf. Comp. Sci. 40 (2000) 71.
- 310 [9] H. C. Tülp, K.-U. Goss, R. P. Schwarzenbach, and K. Fenner, Environ. Sci. Technol. 42 (2008)
311 2034.
- 312 [10] G. Caron, and Ermondi, G. in R. Mannhold (Ed), Molecular Drug Properties: Measurement and
313 Prediction, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2008, pp. 315.
- 314 [11] G. Cruciani, P. Crivori, P. A. Carrupt, and B. Testa, J. Mol. Struct. (Theochem) 503 (2000) 17.
- 315 [12] P. Crivori, G. Cruciani, P. A. Carrupt, and B. Testa, J. Med. Chem. 43 (2000) 2204.
- 316 [13] G. Caron and G. Ermondi, J. Med. Chem. 48 (2005) 3269.
- 317 [14] G. Caron, M. Ravera, and G. Ermondi, Pharm. Res. 28 (2011) 640.
- 318 [15] S. Wold, M. Sjostrom, L. Eriksson, Chemometr. Intell Lab. 58 (2001) 109.
- 319 [16] T. Scior, J. L. Medina-Franco, Q.-T. Do, K. Martínez-Mayorga, J. a Yunes Rojas, and P. Bernard,
320 Curr. Med. Chem. 16 (2009) 4297.
- 321 [17] A. Tropsha, P. Gramatica, and V. K. Gombar, QSAR Comb. Sci. 22 (2003) 69.
- 322 [18] B. Buszewski and S. Noga, Anal. Bioanal. Chem. 402 (2012) 231.

323 **Figure captions**

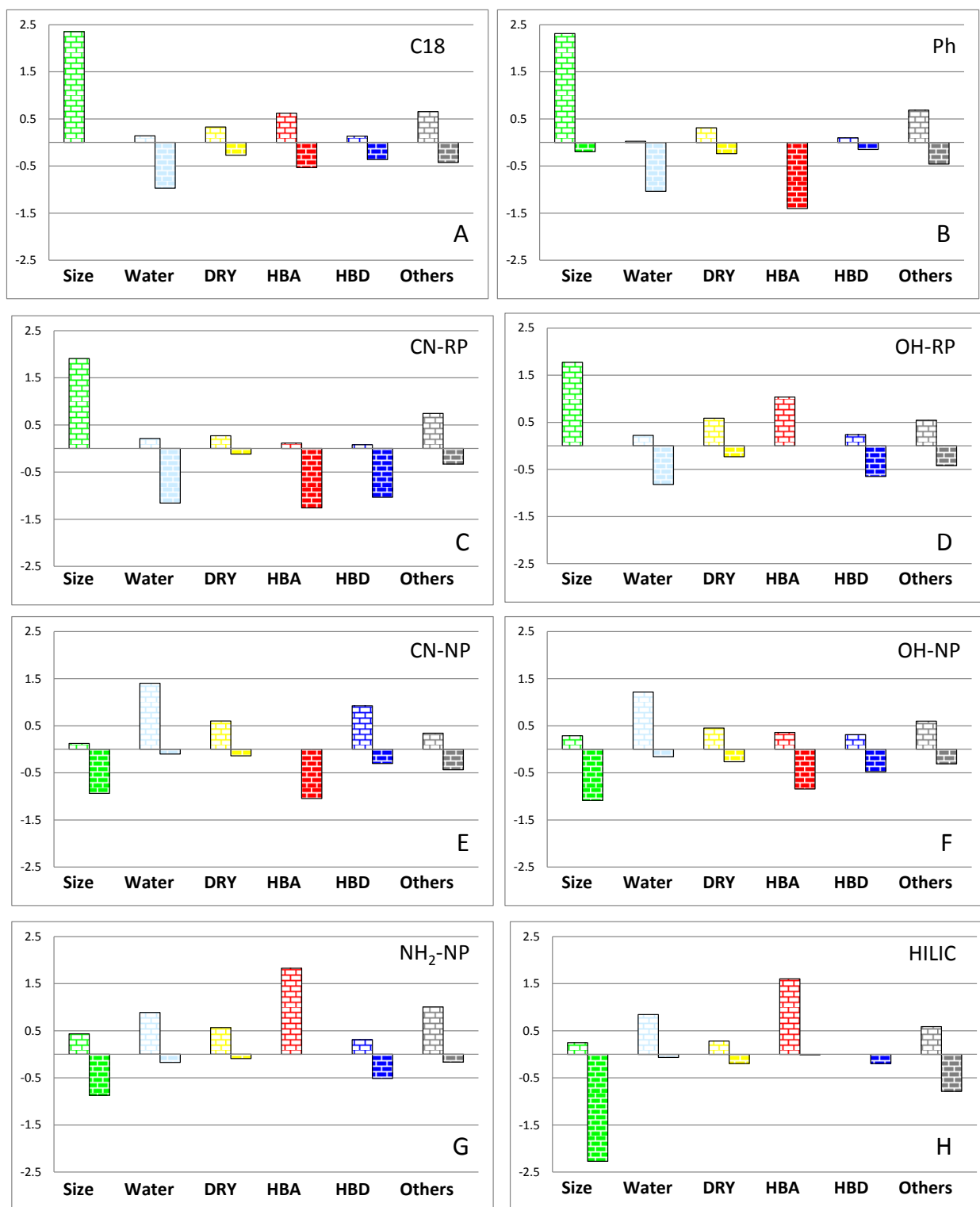
324 Figure 1. Amitrole (a molecule present in the dataset): visual representation of some VolSurf+
 325 descriptors obtained with different probes. A) water (descriptors: W3, W5 and W7) , B) DRY (probe
 326 with hydrophobic properties, descriptors: D1, D4 and D8) and C) HB probes: N1 with hydrogen
 327 bonding donor properties (in the text called HBD for short, color code: blue) forms HB interactions with
 328 solute's hydrogen bonding acceptor (HBA) properties and O with HBA properties (in the text called
 329 HBA for short, color code: red) forms HB interactions with solute's HBD properties.



330

331

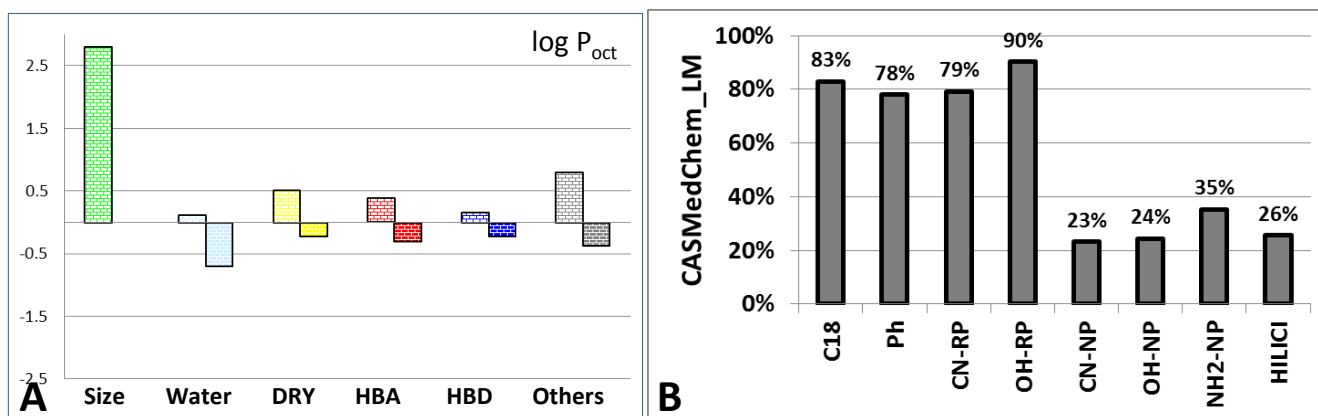
332 Figure 2. A graphical tool to compare chromatographic systems: BR (+) and BR (-) for the eight
 333 investigated systems are reported. Details about the six blocks of descriptors (size, water, DRY, HBA,
 334 HBD, others) are in the text



335

336

337 Figure 3. A) BR (+) and BR (-) for the octanol/water system (data by by Tuelp et al. [9], PLS model: R^2
 338 $= 0.82$, $Q^2 = 0.70$, $LV = 3$, see text for details) and B) LM values to numerically compare octanol/water
 339 and the eight chromatographic systems.



340

341

Table 1. Final PLS models (n = number of observations, R^2 = cumulative determination coefficient, Q^2 = cross-validated correlation coefficient, LV = number of latent variables, RMSE = root mean square of the errors)

System	Abbreviation	Column	Mobile phase	n	R^2	Q^2	LV	RMSE	Missing data	Excluded
RP	C18	YMC-Pack-Pro C18 RS	MeOH/Water 75/25	72	0.72	0.55	3	0.264		c17, c59, c62, c70
	Ph	YMC-Pack-Ph	MeOH/Water 64/36	76	0.80	0.70	3	0.205		none
	CN-RP	YMC-Pack-CN	MeOH/Water 45/55	70	0.70	0.52	3	0.256	c36, c38	c17, c39, c55, c62
	OH-RP	YMC-Pack Diol	MeOH/Water 36/64	69	0.71	0.51	3	0.244	c36, c38	c07, c22, c24, c41, c48
NP	CN-NP	YMC-Pack-CN	n-Hexane/Propan-2-ol/ Di-isopropyl ether 50/40/10	65	0.63	0.46	3	0.214	c07	c09, c10, c16, c19, c59, c60, c66, c70, c74
	OH-NP	YMC-Pack Diol	n-Hexane/Propan-2-ol/ Di-isopropyl ether 55/30/15	69	0.68	0.50	3	0.236	c07	c09, c16, c19, c59, c70, c74
	NH ₂ -NP	YMC-Pack-NH ₂	n-Hexane/Propan-2-ol/ Di-isopropyl ether 55/30/15	70	0.70	0.50	3	0.307	c31	c09, c16, c17, c19, c72
Hydrophilic	HILIC	ZIC-HILIC	AcCN/Water 40/60	69	0.80	0.52	3	0.046	c06, c12, c26, c47, c49, c72, c76	none

Table 2. BlockRelevance (BR) for the eight investigated systems. In parentheses the number of descriptors for the block.

System	Size (7)	Water (19)	DRY (28)	HBA(5)	HBD (5)	Others (18)
C18	2.35	1.11	0.60	0.50	1.15	1.08
Ph	2.51	1.06	0.55	0.25	1.40	1.15
CN-RP	1.91	1.37	0.39	1.11	1.37	1.07
OH-RP	1.77	1.04	0.82	0.89	1.04	0.96
CN-NP	1.06	1.50	0.74	1.23	1.05	0.78
OH-NP	1.37	1.37	0.71	0.79	1.20	0.91
NH ₂ -NP	1.30	1.06	0.66	0.82	1.83	1.17
HILIC	2.52	0.41	0.98	0.20	1.61	1.37

Table 3. LipophilicityMatch (LM) values for the eight systems investigated

	C18	Ph	CN-RP	OH-RP	CN-NP	OH-NP	NH₂-NP	HILIC
C18	100	85	77	82	18	22	33	26
Ph		100	79	76	26	24	26	23
CN-RP			100	79	32	38	41	32
OH-RP				100	26	27	44	29
CN-NP					100	77	66	68
OH-NP						100	70	70
NH₂-NP							100	73
HILIC								100